

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ADMINISTRACAO  
PROGRAMA DE GRADUACAO EM ADMINISTRACAO

PEDRO HENRIQUE FRANZON BENZ

***DATA ANALYTICS APLICADO A PRICING:***  
CONCLUSÕES DA ANÁLISE DA BASE DE VENDAS DE UMA EMPRESA  
MANUFATUREIRA

Porto Alegre

2023

PEDRO HENRIQUE FRANZON BENZ

***DATA ANALYTICS APLICADO A PRICING:***  
CONCLUSÕES DA ANÁLISE DA BASE DE VENDAS DE UMA EMPRESA  
MANUFATUREIRA

Trabalho de conclusão de curso apresentado como requisito para obtenção de graduação na faculdade Administração da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Pablo Cristini Guedes

Porto Alegre

2023

## LISTA DE ILUSTRAÇÕES

Figura 1 - distribuição de equipamentos contratados por dimensionais e capacidade .....	6
Figura 2 - Fluxograma de análise segundo a metodologia CRISP-DM .....	9
Figura 3 - Árvore de decisão do cenário 3 .....	21
Figura 4 - Ranking de importância das variáveis no cenário 1 .....	24
Figura 5 - Ranking de importância das variáveis no cenário 2 .....	25
Figura 6 - Ranking de importância das variáveis no cenário 3 .....	26
Figura 7 - Ranking de importância das variáveis no cenário 3 .....	33

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>3</b>
<b>1.1</b>	<b>DEFINIÇÃO DO OBJETO</b>	<b>4</b>
<b>1.2</b>	<b>JUSTIFICATIVA</b>	<b>5</b>
<b>1.3</b>	<b>OBJETIVOS GERAIS</b>	<b>7</b>
<b>1.4</b>	<b>OBJETIVOS ESPECÍFICOS</b>	<b>7</b>
<b>2</b>	<b>REVISÃO TEÓRICA E COMPREENSÃO DO OBJETO DE ESTUDO</b>	<b>8</b>
<b>3</b>	<b>COOMPREENSÃO E PREPARAÇÃO DOS DADOS</b>	<b>15</b>
<b>4</b>	<b>MODELAGEM</b>	<b>18</b>
<b>4.1</b>	<b>ÁRVORE DE DECISÃO</b>	<b>18</b>
<b>4.2</b>	<b><i>RANDOM FOREST</i></b>	<b>23</b>
<b>4.3</b>	<b><i>EXTREME GRADIENT BOOSTING</i></b>	<b>30</b>
<b>5</b>	<b>RESULTADOS</b>	<b>35</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>39</b>
	<b>REFERÊNCIAS</b>	<b>40</b>

## 1 INTRODUÇÃO

O cenário econômico mundial incerto decorrente da pandemia da COVID-19 obrigou a organizações procurarem por vantagens competitivas. Nesse contexto, um tema que surge com criticidade, dada a repentina alteração de padrões de custos produtivos, demanda e cadeia de suprimento, é o *pricing* (ABDELNOUR *et al*, 2020): a estratégia de precificação de produto vendido ou serviço prestado. Esse contexto, aliado à aceleração de processos produtivos cada vez mais dinâmicos e informatizados (Mckinsey Digital, 2020), gerou uma necessidade por melhor gestão das informações geradas pelas corporações.

Entretanto, por vezes a ausência de monitoramento dessas informações implica a criação de uma base de dados sem a realização de uma análise aprofundada. Neste estudo buscar-se-á trabalhar junto a uma empresa fabricante do ramo de elevadores e soluções de mobilidade urbana para identificar oportunidades em informações ainda não trabalhadas. Especificamente, visa a desenvolver análises em uma base de dados para auxiliar no processo de precificação de propostas comerciais, na previsão de desconto e margem.

A organização atua nos mercados de manufatura de produto e de prestação de serviços e, para crescer sua participação de mercado num cenário econômico desafiador, foi identificada a possibilidade de adoções de novas abordagens no trato com seus clientes. Hoje a companhia não utiliza informações quantitativas para definir diferentes níveis de margem de contratação do produto ou serviço, há apenas uma premissa que um negócio não pode ser fechado com um preço líquido abaixo do custo total planejado<sup>1</sup>. Assim, pretende-se identificar padrões quantitativos e qualitativos nas propostas comerciais para ajudar a definir o desconto adequado a uma negociação.

Em contato com a companhia, percebeu-se que, ao longo dos últimos anos, foi gerada uma vasta base de informações de propostas e de contratos, com dados qualitativos e quantitativos respectivos a cada proposta comercial. Ainda assim, a atual estratégia de precificação da companhia conta com uma abordagem bastante simples, com a definição de uma margem de lucro pré-definida (*mark-up*) em cima do custo do produto, podendo-se

---

1 Ao relacionar receitas e custos no demonstrativo de resultado do exercício, Ross, Westerfeld e Jaffe (2015, 212-215), classificam a margem de contribuição como receita líquida - custo variável, ou seja, qual o lucro resultante da venda, subtraídos todos os materiais utilizados diretamente para a produção do bem ou serviço. A margem bruta obtém-se pela margem de contribuição - custos fixos, ou seja, qual o lucro resultante da receita descontada de todos os gastos diretos para a transformação do objeto ou prestação do serviço. A empresa analisada considera para a margem bruta todos esses custos; quando define que não irá vender com margem bruta negativa, equivale a dizer que o limite inferior ao qual dado vendedor pode negociar uma proposta é o total de custo orçado na fabricação do equipamento e na sua instalação na obra (compra de material, mão de obra, depreciação, logística interna).

conceder um desconto genérico, sem um uso aprofundado das informações disponíveis para influenciar a decisão.

Ao longo do trabalho serão explicados os conceitos do modelo de negócio da empresa, bem como a relação entre as variáveis físicas dos equipamentos, premissas financeiras e outros dados essenciais à análise. Pretende-se verificar quais são as relações que influenciam positivamente na geração de margem da proposta contratada para ajudar a direcionar a criação de valor na relação com o cliente.

## 1.1 DEFINIÇÃO DO OBJETO

Esse estudo tem como objeto uma empresa manufatureira que atua no ramo de produção de elevadores, localizada no Rio Grande do Sul e que tem como principais negócios a transformação do produto, instalação na obra e prestação de serviços de manutenção e reparos. A companhia atua em toda extensão do território nacional e conta com mais de 5 décadas de atuação.

O contexto em que se insere o trabalho é na etapa de compra e venda do equipamento para instalação em uma obra nova. Procura-se analisar uma base de dados correspondente ao histórico de negociações com o objetivo de verificar quais foram as características determinantes do desconto negociado nas propostas que a companhia realizou com seus clientes. Em razão das muitas possíveis características físicas de um elevador, a incidência de dois equipamentos idênticos (muito embora eles possam ser agrupados se consideradas algumas aproximações de variáveis – por exemplo, catalogar todo o pacote elétrico do equipamento em vez de considerar módulos, rabichos e placas individualmente) é rara, e, assim, cada proposta apresenta um nível de desconto diferente no fechamento da proposta. Ao longo dos anos, a companhia refletiu no financiamento do cliente via desconto sua postura de crescimento da presença de mercado: quanto mais agressiva a estratégia, maior era o desconto. Entretanto, devido ao aperto nas condições econômicas em razão da COVID-19, a tática foi revista com um viés de maior atenção à manutenção da margem na contratação.

Com essa restrição, veio à tona a questão de, dada a diversidade de características dos elevadores negociados, quais seriam as negociações mais interessantes para a manutenção da lucratividade da empresa, bem como quais seriam as características físicas desses equipamentos. Outros pontos como localização da obra de instalação, relação com o cliente, ano da proposta, segmento do mercado e do cliente também poderiam influenciar na

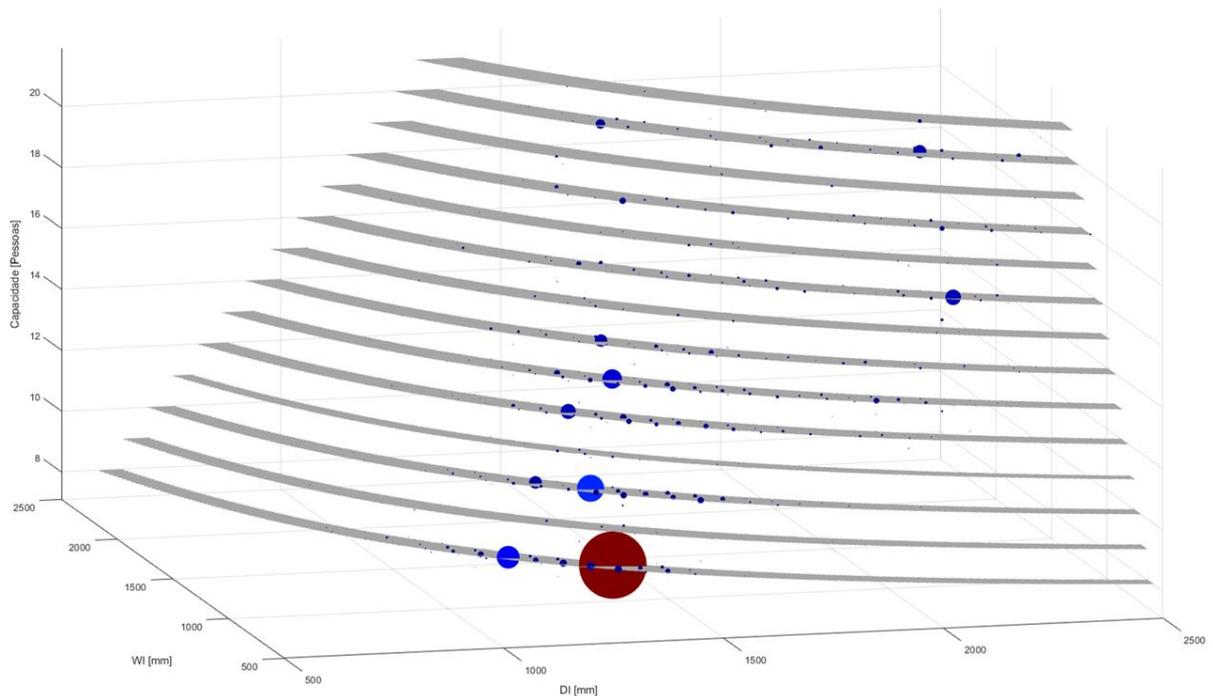
determinação do desconto. Assim, para entender como otimizar a contratação, olhou-se para a base de negociações em busca de uma visualização do problema.

## 1.2 JUSTIFICATIVA

Frente aos desafios que a companhia vem enfrentando, principalmente decorrentes de condições macroeconômicas de mercado e a novas definições corporativas, o estudo realizado pretende auxiliá-la do ponto de vista gerencial a definir uma estratégia comercial de acordo com os vetores que orientam o desconto de uma negociação, focando na geração de margem e crescimento de mercado. Espera-se que as conclusões extraídas poderão auxiliar na definição de estratégias de contratação para a área comercial, bem como darão visibilidade às áreas corporativas das tendências de mercado as quais a empresa está exposta e às áreas de projeto as quais as características físicas que devem ser desenvolvidas para melhor atender às necessidades dos clientes.

Quanto ao portfólio de produtos atual, é disponibilizada uma infinidade de arranjos de propriedades técnicas diversas para cada equipamento. Entretanto, por mais que todo elevador seja de certa forma customizado para se adaptar à obra de instalação, muitas características são padronizadas de acordo com certas definições do produto. Por exemplo, cada intervalo de possibilidade de dimensional de cabina (largura x profundidade) é atrelado à capacidade de carga do equipamento (capacidade). Analisando-se quais as possibilidades de capacidade e dimensionais, pode-se verificar que a maior parte das negociações da empresa ocorrem num intervalo específico dessas características. Para a conclusão que se pretende, é fundamental identificar quais são esses atributos e como influenciam na precificação de uma proposta comercial.

Figura 1 - Distribuição de equipamentos contratados por dimensionais e capacidade



Fonte: o autor (2023).

A precificação de produto e gestão do cliente são questões bastante estratégicas para a empresa analisada, que hoje dispõe de uma metodologia efetiva, mas simples, de *pricing*. O aprofundamento dos critérios de precificação que se procura propor a partir das análises realizadas neste estudo pode dar à organização uma vantagem competitiva na seleção das propostas que efetivamente geram valor na contratação, na formulação de um portfólio direcionado e na estratégia para diferentes mercados regionalizados.

Os resultados serão apresentados ao corpo diretivo da empresa como proposta de geração de valor e, se for o caso, às áreas comerciais como treinamento para utilização da informação na operação. Nesse contexto, os dados e conclusões poderão ser incluídos nos sistemas de gerenciamento de clientes e de vendas, possibilitando aos departamentos de vendas e comercial uma visibilidade melhor das propostas que deverão trazer maior valor a empresa e, assim, direcionar os descontos.

Já do ponto de vista acadêmico, é comum deparar-se com muitos trabalhos em análise de dados como suporte à tomada de decisão gerencial, porém o segmento de elevadores não parece gozar da popularidade que outros tópicos desfrutam. Essa pesquisa é uma contribuição ao desenvolvimento da pesquisa nesse segmento, que guarda inúmeras particularidades, como se verá adiante.

### 1.3 OBJETIVOS GERAIS

Confrontando-se os dados disponíveis, visa-se a investigar qual é o perfil dos contratos que demandam um esforço financeiro da empresa para fechar o negócio e qual o que deixa uma margem mais lucrativa, onde são fechados esses contratos, e quais as características físicas do de equipamento. Assim, a área corporativa terá uma visualização melhor dos vetores de geração de valor nos produtos, bem como a área comercial, na hora de negociar uma proposta, terá um paradigma de desconto para comparação a fim de avaliar se o cliente está exigindo um desconto excepcional ou adequado.

### 1.4 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desse estudo são:

- Verificar quais são os fatores que influenciam no desconto de propostas da empresa;
- Identificar quais são as tendências encontradas nas variáveis que levam à necessidade de conceder maior desconto para fechar uma proposta; e
- Identificar se há a figura do cliente parceiro da empresa e qual o retorno que ele dá em termos de margem.

## 2 REVISÃO TEÓRICA E COMPREENSÃO DO OBJETO DE ESTUDO

Em estudo sobre análise preditiva e árvores de decisão, Chee Sun Lee, Peck Yeng Sharon Cheang e Massoud Moslehpour<sup>2</sup> ressaltam a sua importância no auxílio à tomada de decisão em organizações, em especial para setores produtivos, de logística, de marketing e de vendas, sendo ferramenta capaz de gerar importante vantagem competitiva em termos de qualidade informacional. Citam a *business analytics* como uma área de *data analytics* voltada para negócios e como um arcabouço metodológico qualitativo preciso para extração de conclusões a partir de dados. Investigando as principais tendências do campo, o subdivide em *descriptive analytics* (análise estatística descritiva)<sup>3</sup>, *prescriptive analysis* (análise estatística prescritiva)<sup>4</sup> e *predictive analysis* (análise estatística preditiva)<sup>5</sup>.

Este estudo será abordado pelo viés da *predictive analysis*, dado que, como sustentam os autores, "é o braço de *analytics* que se utiliza de dados, combinações estatísticas e aprendizado de máquina para prever a probabilidade de certo evento acontecer, prever tendências futuras ou resultados com objetivo de melhorar a performance da organização" (LEE *et al*, 2022). Para a implementação de um projeto nesse sentido, ressaltam a importância de adoção de um processo padronizado e propõe um fluxo de sete etapas, começando pelas definições de negócio e projeto até a modelagem dos dados e entrega dos resultados.

Nesse contexto, Rüdiger Wirth e Jochen Hipp<sup>6</sup> propõem uma padronização para o processo de tratamento de dados e métodos de análise de dados. A chamada CRISP-DM (*Cross Industry Standard Process for Data Mining*) surge da necessidade de criação de um quadro de trabalho para o *data analytics*, visando a aumentar confiabilidade, repetibilidade e

---

2 LEE, Chee Sun; CHEANG, Peck Yeng Sharon; MOSLEHPOUR, Massoud. Predictive Analytics in Business Analytics: Decision Tree. **Advances in Decision Sciences**, v. 26, p. 1-30, mar. 2022.

3 Processo de interpretação de conjuntos de dados para identificar certos padrões e tendências. Inclui métodos como o cálculo de médias, desvios padrão, quartis, distribuições de frequência, gráficos, entre outros. A análise estatística descritiva é frequentemente usada como primeira etapa de análise de dados antes da aplicação de técnicas estatísticas inferenciais. É utilizada em aplicações como compreensão de comportamento de consumidor ou tendências de cadeia de suprimento.

4 A análise estatística prescritiva é um agregado de técnicas quantitativas voltado para determinar a ação mais adequada a ser tomada em uma situação específica, baseado em dados e modelos matemáticos. Utiliza técnicas de modelagem para avaliar as consequências de diferentes decisões e identificar a ação que resultará no melhor desfecho ou maior ganho de performance.

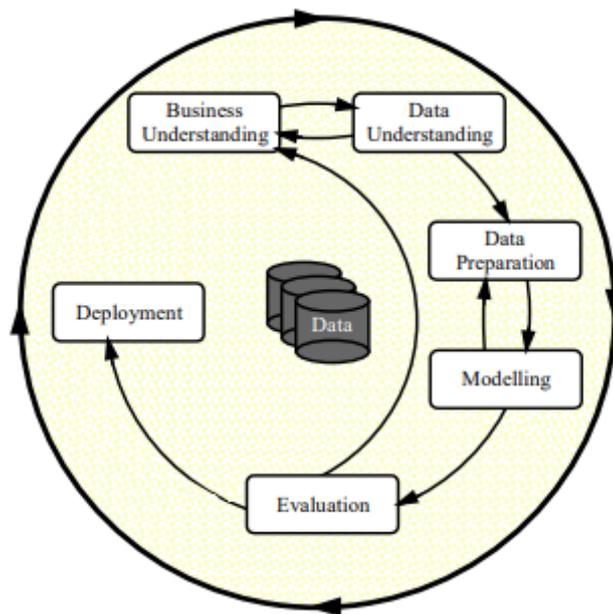
5 Abordagem quantitativa cujo objetivo é usar dados históricos para prever resultados futuros. Ela baseia-se na utilização de modelos matemáticos e estatísticos para estimar o resultado de uma variável dependente a partir de uma ou mais variáveis independentes. Os modelos de análise estatística preditiva incluem, dentre outros, regressão, árvores de decisão, *random forest*. Tais técnicas são amplamente utilizadas para apoiar a tomada de decisão em áreas como finanças, marketing ou saúde.

6 WIRTH, Rudiger; HIPPE, Jochen. CRISP-DM: Towards a Standard Process Model for Data Mining. **Practical application of knowledge discovery and data mining**, p. 29-40, 2000.

gestão de trabalhos. Assim, a abordagem proposta nesse estudo será inspirada na proposta dos autores.

A metodologia sugerida divide o trabalho em 6 fases distintas, mas interconectadas, a saber: compreensão do negócio,<sup>7</sup> compreensão dos dados,<sup>8</sup> preparação dos dados<sup>9</sup>, modelagem<sup>10</sup>, avaliação<sup>11</sup> e entrega<sup>12</sup>. O encadeamento lógico dentre essas etapas é permeado de pontos em que se pode voltar a etapas anteriores, por exemplo, pode ser necessário reavaliar o entendimento de negócio de acordo com as avaliações das conclusões extraídas de um modelo preliminar montado nos dados do negócio.

Figura 2 - Fluxograma de análise segundo a metodologia CRISP-DM



Fonte: Wirth e Hipp (2000).

Assim, mesmo que já pincelados alguns aspectos introdutórios sobre a empresa e a motivação do presente escrito, ainda restam definir algumas premissas de negócio para a

7 Essa etapa consiste em entender o objetivo do projeto, bem como regras e requisitos do negócio

8 Nesse momento é realizado o mapeamento preliminar dos dados necessários e disponíveis para o modelo, bem como uma identificação inicial de possíveis problemas ou formas de análise do conjunto de dados.

9 A fase de preparação dos dados cobre todas as atividades voltadas à construção final do conjunto de dados que será utilizado para alimentar os modelos. Consiste basicamente em coletar, limpar, transformar e validar os dados obtidos.

10 É na modelagem em que são selecionadas e aplicadas as técnicas de análise e configurados os parâmetros adequados a partir do conjunto de dados obtidos da fase anterior.

11 Nesse estágio, já contando com um ou mais modelos consistentes, pode-se partir para testagem da aplicação do resultado nas hipóteses e regras de negócio definidas para a análise.

12 Após geradas todas as análises e mapeados os resultados, passa-se à entrega do material ao cliente da informação, que deverá recebê-lo no formato adequado, com clareza e conveniência das informações.

melhor compreensão das conclusões. Após, a base de dados obtida junto à empresa, contendo todas as variáveis trabalhadas, será exposta e analisada para, então, se comentar as ferramentas analíticas que foram utilizadas para a preparação e modelagem dos dados. Acabadas essas etapas, serão traçadas as conclusões e preparação da entrega do material desenvolvido.

Precificar um produto ou serviço é uma tarefa essencial a atividade empresarial (WU *et al*, 2022), e muitos são os caminhos para se quantificar, em valor monetário, o objeto do negócio. Para uma transação ocorrer, é necessária uma definição para as partes envolvidas acerca no preço da utilidade prestada, sendo ele, idealmente, o maior esforço financeiro que o cliente se proporia a fazer e o menor pelo qual a empresa estaria disposta a prestar o serviço ou entregar o bem (YE *et al*, 2020). No entanto, pode haver transigências quanto a esses limites – um comprador de um produto *premium* pode estar disposto a pagar a mais por algum aspecto intangível do bem, ou o vendedor pode estar inclinado a aumentar o desconto aumentar o volume de vendas e, assim, crescer a sua presença de mercado.

A adoção de uma política de precificação está diretamente vinculada aos objetivos da empresa e à capacidade de geração de valor (Institute of Management Accountants, 2019). Se objetiva-se preservar as margens de lucratividade, pode-se adotar uma estratégia de *cost-plus pricing*<sup>13</sup>. Por outro lado, se o objetivo é aumentar a presença de mercado, o *penetration pricing*<sup>14</sup> pode ser o caminho a ser seguido. Se a intenção é de segmentar a base de clientes, uma estratégia de *value based pricing*<sup>15</sup> pode ser mais adequada.

Nesse contexto, uma estratégia de precificação, para estar alinhada com os objetivos da empresa, necessita ser discutida em múltiplos níveis organizacionais, como sustentam Bonnachi e Perego (2019, p. 13) em seu livro *Customer Accounting*. Os autores propõem, com o aumento da complexidade e da diversidade de produtos e serviços, a adoção de novas políticas de precificação, com viés preditivo, orientadas por modelagem da dados. Referenciam, ainda, que para aferição do sucesso de implementação dessas novas estratégias, é preciso também estabelecer algumas métricas de medição de performance. Olhando para o lado da relação empresa-cliente, sugerem como elementos chave nesse monitoramento, o

---

13 Estratégia de acréscimo de uma margem pré-definida em cima do custo estimado do produto ou serviço, privilegiando a retenção da lucratividade em detrimento de um potencial crescimento de mercado.

14 Prática de definir um preço abaixo do praticado e do orçado para crescer a presença de mercado, resultando na perda de lucratividade.

15 Estratégia de definição do preço de acordo com o valor percebido pelo cliente, pode apresentar margens diferentes de acordo com os segmentos.

*customer profitability*<sup>16</sup>, o *customer lifetime value*<sup>17</sup> e o *customer equity*<sup>18</sup>. Este último pode ser utilizado como uma aproximação do valor da empresa, sendo muito útil para os objetivos da administração financeira, conforme escrevem Silveira, Oliveira e Luce (2012, p. 1.753).

Em artigo publicado na *Journal of Business Research*, Luce, Silveira e Heldt (2021, p. 1-4) também propõem a adoção de métricas prospectivas de projeção de rentabilidade em detrimento de estratégias baseadas em dados passados. Para chegar ao já comentado *customer equity*, sugerem a sua avaliação por um modelo RFM/P<sup>19</sup>, com foco tanto no cliente como no portfólio de produtos. Argumentam que a adoção desse modelo bifocado permite à empresa ter uma visão mais apurada de sua rentabilidade futura, bem como intuições mais precisas acerca do valor do seu portfólio de produtos para cada segmento de cliente. Em linha com essa perspectiva, Bonnachi e Perego (2019, p. 16) suportam a utilização do *customer lifetime value* como a ferramenta de análise prospectiva de rentabilidade por cliente mais adequada, dado que atuaria como o limite superior do que a firma deveria estar disposta a dispende para adquirir ou reter um cliente.

Entretanto, definir esse limite pode não ser tão simples no caso analisado – algumas peculiaridades permeiam o mercado em que atua a empresa estudada. Quando se fala do ciclo de vida de um elevador, podemos dividi-lo em três etapas: instalação, manutenção e modernização. A instalação ocorre quando uma obra recebe pela primeira vez um equipamento, enquanto a manutenção começa quando a primeira fase é concluída, em que é prestado um serviço de revisão periódica e troca de peças pontuais. Quando é necessária uma atualização mais profunda do equipamento pelo ingresso de novas tecnologias, ou por avarias e desgaste, começa a terceira fase do ciclo de vida: a modernização, em que podem ser trocados alguns componentes (cabos, portas de pavimento, máquina de tração, quadro de comando...) ou o equipamento por inteiro. Assim, para se analisar o *lifetime value* de um elevador, é necessário endereçar esse ciclo como um todo.

---

16 Diferença entre receita recebida e custos incorridos de todos os produtos na relação com um determinado cliente durante um período específico.

17 Valor dos fluxos de caixa futuros projetados para a relação com determinado cliente trazidos a valor presente líquido, descontados pelo custo médio ponderado de capital da empresa.

18 Soma dos valores de fluxos de caixa futuros para toda a base de clientes atuais e potenciais da empresa. Autores como Luce, Schmitt e Heldt (LUCE, Fernando Bins; SILVEIRA, Cleo Schmitt; HELDT, Rodrigo. Predicting customer value per product: From RFM to RFM/P. *Journal of Business Research*, [s. l.], v. 127, p. 444-453, abr. 2021, p. 8). Defendem o seu uso (*Customer Equity share*) na definição do tamanho de mercado da organização em detrimento do *market share*.

19 Método de análise de *customer value* orientado por 3 fatores: atualidade das interações com o cliente, frequência das transações e seus valores. Os autores aqui adicionaram uma dimensão de análise por segmento de produto ou serviço, dado que os clientes nem sempre consomem os mesmos produtos.

Além da complexidade do ciclo de vida do equipamento, a relação com o cliente também não é o padrão que se espera da venda de varejo – para exemplificar, uma das nuances é a depender do estágio do ciclo de vida, o mercado alvo pode mudar. Um elevador novo, via de regra, será instalado num prédio em construção, enquanto a manutenção e a modernização podem incidir em qualquer equipamento já existente, da mesma marca ou de outras. Assim, têm-se relações comerciais com clientes fundo de investimento imobiliário, incorporadoras, construtoras, administradoras de imóveis, condomínios residenciais, entre outros. Essa heterogeneidade de clientes é outro desafio para a análise do *customer lifetime value*.

O escopo desse trabalho será analisar a primeira parte do ciclo de vida do elevador, a instalação, pois considerações sobre o *lifetime value* do equipamento considerando a sua integridade necessitaria tangenciar variáveis mais complexas e das quais a empresa não tem controle, como durabilidade, condições de uso e cuidado por parte do cliente. Outro ponto importante é a relação entre empresa e cliente e o seu valor ao longo do tempo. Considerando-se cada proposta de novo equipamento no contexto da relação da empresa com o cliente, também se está integrando variáveis de difícil mensuração, pois depende muito fortemente da relação entre os representantes comerciais – um mesmo cliente pode optar, por exemplo, por contratar um equipamento para ser entregue no Ceará pelo representante da divisão de São Paulo por já terem experiências passadas.

Dadas essas restrições, a análise se concentrará especificamente naquela primeira fase do ciclo de vida do elevador, a instalação, e numa perspectiva *standalone* de propostas comerciais. Serão investigados que fatores são determinantes para que dada proposta, por si só, traga maior ou menor margem para a empresa, sem olhar para relação com clientes e outras etapas do ciclo, e se é possível a cada nova proposta, estimar uma base de desconto e uma margem para que o vendedor possa ter como referência a partir de características técnicas dos equipamentos ou das condições de local.

A proposta que se faz de uma precificação dinâmica é diferente da atual formulação de preço que a empresa utiliza, o *cost plus*<sup>20</sup>. Alterações na forma de aferição de preço para o produto não é algo simples e que pode ter consequências na relação com os clientes. David Prakash e Martin Spann<sup>21</sup> observam o efeito que um *pricing* dinâmico tem em relação ao

---

20 O método *cost plus* é nada mais que o custo orçado acrescido de uma margem de lucratividade esperada. A empresa analisada utiliza essa abordagem independentemente de qualquer condição ou premissa de característica da proposta, do equipamento, do cliente ou estimativas de valor estratégico da negociação.

21 PRAKASH, David; SPANN, Martin. Dynamic pricing and reference price effects. **Journal of Business Research**, vol. 152, p. 300-314, 2022.

preço de referência pelo qual o mercado percebe o valor do produto. Em estudo publicado na *International Journal of Production Economics*, Meng Wu, Yun Ran e Stuart Zhu<sup>22</sup> exploraram a relação de diferentes métodos de precificação de acordo com condições de mercado e características de clientes.

Nesse contexto, ganha importância a definição de um sistema adequado de precificação para o negócio. Como auxílio a essa tomada de decisão, o já citado estudo de Chee Sun Lee, Peck Yeng Sharon Cheang e Massoud Moslehpour ressalta a importância da técnica de árvore de decisão no campo de *predictive analysis*. É um método que ganhou popularidade em aplicações de análise estatística de dados, particularmente em aplicações de *business analytics*, devido a sua simplicidade – a partir de uma série de perguntas sobre as características dos dados, divide um conjunto de dados em subconjuntos menores, com o objetivo de encontrar uma estrutura que permita fazer previsões precisas sobre o resultado de interesse com base nas características dos dados (LEE *et al*, 2022).

Essa técnica pode ser utilizada para avaliar tanto variáveis numéricas, quanto categóricas, e o processo de mensuração difere de acordo com essa natureza: quando se analisa variáveis categóricas, chama-se modelo de classificação, enquanto para numéricas, de regressão. Para aquelas, a árvore de decisão escolhe o dado de entrada que melhor divide o conjunto de dados em subconjuntos com valores da variável de saída semelhantes. Já no caso dessas, o modelo escolhe o ponto de divisão que melhor separa o conjunto de dados em subconjuntos com valores da variável de saída semelhantes.

O processo de construção da árvore de decisão é repetido recursivamente para cada subconjunto até que seja atingido um critério de parada, como o número de nós determinado no modelo. Ao final, a estrutura resultante pode ser utilizada para se fazer previsões para novos dados baseados nas perguntas feitas na árvore.

Essa técnica é utilizada em aplicações como prever o preço de *commodities* (LIU, *et al*, 2020), ou criptomoedas (RATHAN *et al*, 2022). Entretanto, alguns obstáculos podem ser encontrados na sua aplicação (ZHA *et al*, 2022), como problemas de interpretação no caso de variáveis categóricas com muitas categorias, *overfitting* em modelos muito complexos e maior volatilidade, dado que as variáveis independentes para modelar a análise são escolhidas aleatoriamente.

Uma alternativa à utilização da árvore de decisão é a técnica de *random forest*, um algoritmo de aprendizado de máquina que consiste na construção de várias árvores de decisão

---

22 WU, Meng; RAN, Yun; ZHU, Stuart X. Optimal pricing strategy: How to sell to strategic consumers? *International Journal of Production Economics*, v. 244, p. 1-16, 2022. p. 4.

e combina as previsões resultantes para melhorar a precisão do modelo (LARIVIÉRE *et al.*, 2005). Ao invés de construir uma única estrutura, são modeladas  $n$  árvores em subconjuntos aleatórios do conjunto de dados, com cada uma fazendo suas próprias previsões e a conclusão, no caso de modelagem de classificação, sendo adotada por votação e no de regressão, por média.

A técnica é utilizada numa gama similar de aplicações: Budak e Sarvari (2021, p. 4) sugerem uma abordagem estatística por meio da resolução de *random forests* para determinar a otimização da margem de serviço de frete. Adentuji, Akande, Ajala, Oyewo, Akande e Oluwadara<sup>23</sup> se valeram do método para propor um modelo de predição do preço de moradias individuais na área urbana de Boston, Estados Unidos, a partir de fatores físicos.

Outros métodos poderiam ainda ser utilizados, como o *gradient boosting*<sup>24</sup> – técnica utilizada em aplicações baseadas em regressão, similar à estrutura de árvore de decisão, mas menos vulnerável ao *overfitting*<sup>25</sup>. Nessa linha, para o presente estudo, a fim de determinar qual deverá ser o limite de desconto adequado a uma nova proposta, propõe-se uma análise por um algoritmo de *random forest* para resolver uma regressão, modelada com o desconto como variável alvo e uma série de *features* como variáveis independentes.

---

23 ADENTUJI, Abigail Bola *et allia*. House Price Prediction using Random Forest Machine Learning Technique. **Procedia Computer Science**, vol. 199, p. 806-813, 2022.

24 Técnica de aprendizado de máquina que utiliza vários modelos de classificação ou regressão para produzir uma única previsão combinada. Para gerar esse resultado, utiliza de um método chamado *gradient descent*, que, visando a encontrar o mínimo global de uma função de perda, ajusta a cada iteração da sua execução o modelo testado. Um dos principais benefícios em relação a métodos como o *random forest* é a capacidade de lidar com *overfitting* em modelos com variáveis de muitas dimensões.

25 Fenômeno que pode ocorrer em análises preditivas em que os dados se ajustem ao subconjunto de treinamento mas não se generaliza bem a novos dados. Isso pode ocorrer num modelo muito complexo, com muitas variáveis, ou quando há poucas observações de dados no conjunto analisado.

### 3 COOMPREENSÃO E PREPARAÇÃO DOS DADOS

Os dados obtidos junto a empresa vieram no formato de uma tabela, cujas linhas representam cada elevador e as colunas as suas características. Existem vários tipos diferentes de características: financeiras, físicas do equipamento, dados quanto ao cliente, quanto ao edifício da obra e quanto à filial que registrou o contrato. Ao total, coletou-se mais de 220.000 observações, registrando 195 características específicas.

O conjunto de dados contém observações que datam de janeiro de 2012 até setembro de 2022, negociados com clientes privados ou públicos, para toda e qualquer tipo de obra. Ocorre que nem toda negociação foi fechada, e isso é fácil de perceber na base de dados, dado que as propostas perdidas não contêm dados de segmentação de cliente ou mercado.

Assim, já foi possível fazer um primeiro filtro, excluindo-se as propostas perdidas. Isso foi feito por dois motivos, dado que além da falta de informações de algumas características, não há informação de o porquê a negociação fracassou (preço mais alto, produto não atendia à especificação técnica, por exemplo), então analisar o perfil de desconto dessas negociações pode acrescentar uma incerteza no modelo.

Para o modelo, também foram excluídas todas as propostas relativas a órgãos públicos da base, uma vez que a forma de contratação difere muito das negociações comerciais padrões – o valor da proposta tem muito mais vinculação com a dotação orçamentária que o contratante dispõe no edital de licitação do que com uma concorrência por preço, o que polui a compreensão do desconto necessário para fechar a proposta nesses casos. Pelo mesmo motivo também foram removidos contratos em que a empresa participou como parte de um consórcio ou obras de infraestrutura, como metrô e aeroportos.

Os dados filtrados representaram uma redução para 15.279 observações. Após essa etapa inicial, passou-se a seleção de quais características seriam interessantes e quais seriam dispensáveis para as conclusões. Assim, inicialmente, optou-se por limpar a seção dados financeiros<sup>26</sup> que não fossem relativos ao desconto, tais como custos de frete, encargos

---

26 Ao todo, a seção de dados financeiros compreende 54 colunas com informações relativas ao custo e preço do equipamento contratado. Dado que o modelo de negócio da empresa considera uma abordagem custo + margem para a definição do preço, todos os dados relativos a preço são correlacionados ao desconto. Optou-se pela exclusão desses dados em para evitar algum provável viés dessas informações com a variável alvo da análise. Para entender melhor esses motivos, além das explicações já observadas, é necessário entrar em detalhe na estratégia de precificação adotada pela companhia estudada: para a empresa analisada, o preço do produto é definido a partir da consolidação do custo, acrescido de uma margem, e em cima dessa composição é definido o preço de lista sob o qual incide o desconto. O custo da proposta é oriundo de três fatores, o frete (que varia em função de localidade, modal e ocupação de veículo transportador), a instalação (que varia de acordo com o custo de mão-de-obra considerado no local de instalação e a quantidade de horas requisitadas para a colocação do

tributários, entre outros. Também foi reduzida a quantidade de opcionais mostrados na base<sup>27</sup>. De maneira geral, essas etapas representaram o corte de 151 colunas de dados.

Essas medidas foram adotadas por entendermos que tanto nem todo fator é determinante na hora de auferir o desconto da proposta, quanto muitos dados trariam redundância para a análise. Quanto as características físicas, mesmo reconhecendo que são cruciais na definição de o que é um elevador, algumas são mais relevantes que outras, como por exemplo Linha e Cabina. A Linha define a plataforma comercial pela qual o cliente conhece o produto no catálogo, sendo essencial na relação e comunicação comercial, e a Cabina incorpora a maior parte da experiência de usuário na utilização do elevador. Ambos os fatores automaticamente fixam outros parâmetros no que diz respeito a características técnicas e físicas do equipamento: uma dada cabina, para uma certa linha, pode ter um dimensional único (por exemplo, para a linha Synergy 100 S, a Cabina e os Dimensionais de Cabina são fixados), e essa sobreposição poderia criar um viés indesejado na análise. O mesmo aconteceria com certos dados financeiros, como *Desconto* e *Margem Bruta*, pois como visto, a última é definida em função do primeiro. E como o objetivo da análise do estudo é analisar o *Desconto*, optou-se por excluir *Margem Bruta* e outros dados relativos a preço e custo que pudessem enviesar o modelo.

Num último esforço de tratamento dos dados, para analisar questões relativas ao tratamento com o cliente e de segmentação de mercado, alguns dados foram aglutinados, como Acabamento do Prédio (Funcional, *Premium* ou *Comfort*) e Destinação (Residencial ou Comercial). Também foi criada uma coluna chamada “contratos\_por\_cliente”, que soma quantos contratos fechados o cliente de uma proposta tem no total da base, para identificar se existe alguma forma de cliente parceiro e se essa relação poderia beneficiar a empresa no tocante à margem da negociação. Por fim, existem duas hierarquias em que são divididas as unidades da ponta comercial da companhia: filial e divisão. O nível filial apresenta maior granularidade, motivo pelo qual se optou pela sua remoção, pois o nível divisão sintetiza com

---

equipamento) e material (que é definido a partir de uma lista de materiais necessários para a montagem e varia de acordo com as características técnicas do elevador).

27 A definição técnica das características do elevador se dá pela escolha de opcionais, por meio dos quais o sistema de venda (E-sales) informa o Engineering (sistema de especificação de materiais) e recebe deste último a lista técnica de peças para a construção do equipamento. Por exemplo, uma definição importante para uma obra é a existência de uma casa de máquinas na obra a ser instalado o equipamento. Quando há casa de máquinas e a velocidade configurada é até 60 metros por minuto (1 metro por segundo), o E-sales automaticamente informará ao Engineering a necessidade do opcional 150 – Linha Synergy Sem Casa de Máquinas. Com mais de 400 opcionais possíveis, existe um arranjo praticamente infinito de maneiras diferentes de se configurar um equipamento, e ao longo dos anos, novas definições de engenharia alteraram os números de opcionais que desempenhariam papel equivalente na hora da especificação dos materiais na lista técnica. Por esses motivos, optou-se por limpar também da base analisada todas as informações referentes aos opcionais (97 colunas), posto que também seria possível analisar as características físicas dos elevadores em outros vetores.

menos complexidade as condições de mercado em que a unidade da empresa está inserida – por exemplo, as filiais Santa Maria e Caxias do Sul são duas das filiais da divisão RS Interior –, o que, por outro lado, não impede de caso se trate de um mercado muito exclusivo, a divisão equivalha exatamente à filial, como é o caso de São Paulo, em que a divisão SP Capital tem apenas uma filial.

## 4 MODELAGEM

Uma vez gerada a base de dados a ser analisada, com todas as seleções e filtros propostos, terminou-se com um conjunto de 15.279 linhas por 20 colunas, contendo informações de propostas comerciais fechadas, com clientes, pessoa física ou pessoa jurídica de direito privado, entre os anos de 2012 e 2022, para todas as localidades do Brasil, cujos equipamentos foram contratados para operar em edifícios com destinação comercial ou residencial, de todos os tipos de acabamento. Para então se operar as análises destinadas a cumprir com os objetivos desse estudo, foi idealizada uma bateria de três testes, com diferentes metodologias, para se explicar a determinação do desconto de uma proposta comercial.

No escopo desse estudo, realizou-se uma análise de regressão utilizando três diferentes algoritmos: árvore de decisão, *random forest* e *gradient boosting*. A árvore de decisão é uma técnica de modelagem preditiva que utiliza uma estrutura em forma de árvore para representar relações hierárquicas entre variáveis preditoras e uma variável resposta. Já o *random forest* é um algoritmo de aprendizado de máquina baseado em árvores de decisão, onde várias árvores são construídas a partir de diferentes subconjuntos de dados e variáveis. Por fim, o *gradient boosting* é um método de aprendizado de máquina que combina vários modelos simples de árvore de decisão para criar um modelo mais complexo e preciso.

O objetivo da aplicação desses métodos foi o de analisar a relação entre as características dos elevadores e o desconto oferecido nas negociações. A escolha de diferentes algoritmos de regressão se deu em função de suas características específicas e habilidades na modelagem de dados complexos, permitindo que diferentes perspectivas fossem utilizadas na análise. Ao final do processo, foram obtidos modelos que permitiram a identificação das variáveis mais importantes para a definição do desconto, bem como a avaliação de sua influência em novas negociações. Também pôde-se visualizar potenciais pontos de melhoria na definição dos cenários e espaço para novas análises de cenários, por exemplo como se comporta a base apenas para uma Divisão, ou para um determinado segmento de mercado ou linha de produto.

### 4.1 ÁRVORE DE DECISÃO

O primeiro modelo testado foi a árvore de decisão, devido a sua simplicidade de implementação e de interpretação das conclusões. O cenário contemplou o conjunto de dados

após todos os filtros e seleções descritos. Os parâmetros escolhidos foram os padrões do pacote *rpart*<sup>28</sup> no R: todas as variáveis foram consideradas com o mesmo peso em seis nós de decisão.

Como resultado, foi verificado o ordenamento das seis variáveis mais determinantes ao modelo, da mais relevante à menos: Modelo (36,4), Cabina (9,4), Velocidade (3,8), Divisão (3,7) e Percurso (2,6) e paradas (2,4). Da análise da árvore, o ponto de corte de desconto para a variável Modelo, no primeiro nó, foi 35%, e se a avaliação fosse diferente de 15, 16, 23, 25, 26, 27, 28, 29, 30, 31, 34 e 35, ele sobe para 45%, e se for igual, cai para 34%. Nesse nó, que dividiu a amostra em dois subconjuntos que totalizam, respectivamente, 12% e 88% dos dados iniciais, pôde ser notada uma diferença muito grande no desconto. Isso se deve ao fato de que os equipamentos em que foram verificados modelos diferentes dos listados, também foram negociados em anos em que a política da empresa era agressiva no desconto, mas não foram negociados em anos com uma postura mais conservadora na captação de novas obras. Por exemplo, o modelo 29 corresponde à plataforma elétrica de grande parte das linhas Synergy 100 e FDN, que foi negociada ao longo de todo o período de análise. Já os 12% que foram segregados foram, em sua totalidade, negociados até o ano de 2015.

Dessa análise preliminar, constatou-se uma diferença significativa no Modelo<sup>29</sup> (plataforma elétrica) que ocasionou o primeiro nó: os equipamentos segregados nos 12% foram obsoletados pela companhia. Isso ocorre por alterações de tecnologia, mas não implica necessariamente numa diferenciação do produto. Dessa forma, optou-se por equiparar os diversos modelos aos seus equivalentes atuais para reduzir o ruído ocasionado por questões de necessidade técnica e que não trariam diferenciação do equipamento em si. Nenhum outro filtro, por hora, foi cogitado em relação as propostas do outro subconjunto de dados.

Com o conjunto de dados ajustado para o ajuste das observações de Modelo citadas, o novo modelo apresentou como variáveis mais influentes Acabamento (10,6), Cabina (6,6), Linha (5,5), Building\_type (3,1), Divisão (3,0) e Mercado (0,2). Válido notar que na execução

---

28 O *rpart* é um pacote de R para construção de árvores de decisão. Ele usa um algoritmo de particionamento recursivo que divide a população em grupos menores e menores com base nas variáveis explicativas. A função básica para construção de árvores de decisão é a “*rpart()*”, com vários parâmetros que podem ser configurados para personalizar a modelagem. No caso do cenário analisado, não houve qualquer alteração nos parâmetros padrão de execução ou de controle da árvore.

29 O modelo é um atributo definido na especificação das peças de um elevador em que consta informações respectivas ao quadro de comando, como por exemplo forma de montagem, se a unidade de potência é dentro ou fora do quadro, se há necessidade de inversor de potência. No entanto, alguns produtos são definidos pelo modelo, quando dois elevadores de mesma linha e têm plataformas elétricas diferentes para atender a uma norma técnica.

anterior, o RMSE apresentado havia sido marginalmente menor que o novo – 0,075 em relação aos atuais 0,078. Dessa vez, o primeiro nó, com desconto médio de 34%, seccionou os dados em dois subconjuntos, um com as cabinas One e Art Collection, cujo desconto médio foi de 19% e representou cerca de 3% do total da amostra. Já os outros 97% apresentaram um desconto médio de 34%, para todas as outras cabinas.

Em uma avaliação focada no subconjunto de dados formado pelas cabinas One e Art Collection, verificou-se que as propostas foram negociadas entre 2015 e 2020. No que diz respeito as vendas relativas à cabina Art Collection, somam 20 equipamentos, sendo que 14 foram negociados num único contrato, fechado com um desconto de 24% (bastante inferior à média da amostra de 34%). Já os elevadores da cabina One apresentam as mesmas características de modelo, linha, velocidade, capacidade e limite de altura que a linha Synergy 100 S e pararam de ser negociados com o lançamento do Synergy 100 e Synergy 100 S. Houve uma alteração do produto no ponto de vista da plataforma comercial, e não simplesmente a exclusão da estrutura técnica de um produto. Importante frisar que da mesma forma que o Synergy 100 S é uma tentativa de padronização do produto Synergy 100 e, que por um ganho de escala projetado, a empresa considera um padrão de custo mais baixo para aqueles equipamentos, o Synergy One teve a mesma concepção de desenvolvimento de produto – altamente padronizado e, assim, poderia ser oferecido a um desconto “natural”.

Com esse viés, novamente optou-se por uma adaptação do cenário. Da mesma forma, optou-se pelo ajuste da base para que o Synergy One fosse absorvido pelo produto que o tornou obsoleto em termos de portfólio, mas que continuou com a sua proposta de valor comercial. Assim, dado que os equipamentos com proposta de valor similar aos de cabina One não pararam de ser negociados, mas foram substituídos por outro produto, optou-se pela sua permanência na análise. Interessante notar que o One surgiu como uma alternativa de menor preço aos produtos da companhia e, quando implementado, pôde-se verificar um impacto na margem de contratação e no desconto dessas negociações<sup>30</sup>. Entretanto, à medida em que o mercado absorveu o produto, o desconto aumentou com a convergência do preço de referência em relação ao novo preço. Durante a pandemia da COVID-19 o produto entrou em *phase-out*, dando lugar ao Synergy 100 e Synergy 100 S e, talvez por uma situação atípica de mercado, mesmo com uma política de desconto mais conservadora, a empresa não conseguiu

---

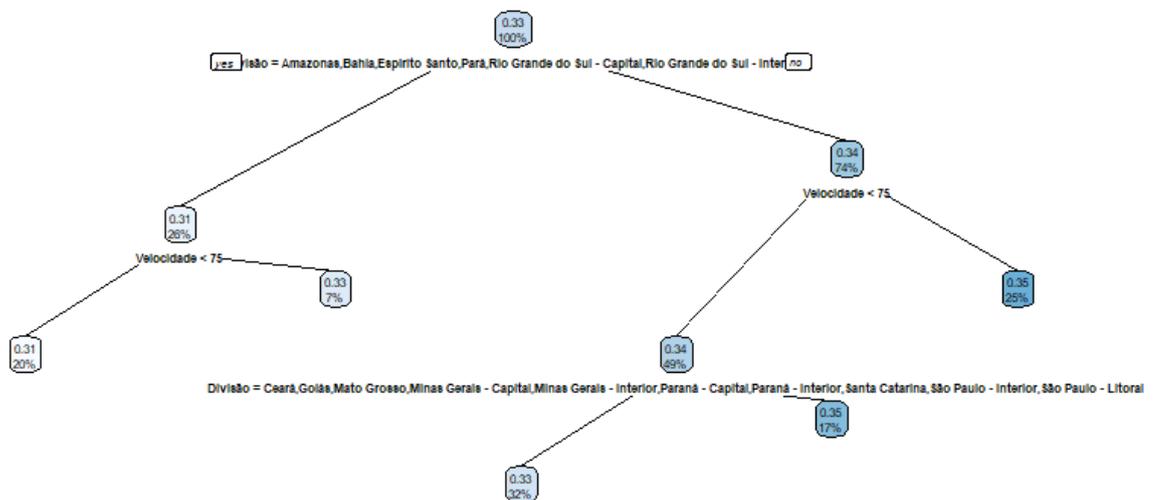
30 Prakash e Spann (PRAKASH, David; SPANN, Martin. Dynamic pricing and reference price effects. **Journal of Business Research**, vol. 152, p. 300-314, 2022, p. 4) argumentam que o consumidor racional, quando se depara com um preço diferente do preço-referência, tende a adquirir o produto quando está abaixo e a não efetivar a negociação caso esteja acima. Entretanto, sendo o preço de referência um consenso tácito entre os participantes do mercado, à medida que o este absorve o novo preço, o preço-referência converge para esse novo preço.

repetir o sucesso inicial do seu produto popular, com o Synergy 100 S exigindo um desconto de 33% no preço de lista, em média para fechar uma negociação em 2020, já quando foi lançado.

Quando o cenário foi modelado novamente, verificou-se que o nó seccionou os dados de acordo com as Cabinas P4000, New Art Collection e KM18. Essa última é a Cabina aplicada ao Synergy 100 e Synergy 100 S, que acabaram absorvendo na base os dados do Synergy One a partir do ajuste. Os equipamentos com essas Cabinas apresentaram um Desconto médio de 33% e representam 21,7% do total da base.

Para tentar mitigar o impacto do fator temporal, propôs-se executar um novo cenário, dessa vez com um filtro de Ano maior que 2015. A quantidade de observações nesse novo cenário foi reduzida para 7.895 e as variáveis de maior importância foram: Divisão (0,9), Velocidade (0,3), Percurso (0,2), Paradas (0,2), Capacidade (0,1) e Cabina (0,1). Novamente, a cabina One apareceu em destaque, totalizando 108 equipamentos, junto com uma unidade de Montacarga<sup>31</sup>.

Figura 3 - Árvore de decisão do cenário 3



Fonte: o autor (2023).

Nesse cenário, como esperado, com a exclusão de observações de datas anteriores a 2016, houve uma redução do desconto praticado (33,4% contra 37,3%). Com a base padronizada em relação à sucessão de Modelos e produtos obsoletos, o modelo utilizou a

31 Elevador pequeno geralmente utilizado para transitar pequenos objetos entre andares, com aplicação comercial ou industrial, como pratos em um restaurante ou utensílios em um hospital

Divisão para fazer a secção dos dados – se a proposta foi negociada em Amazonas, Bahia, Espírito Santo, Pará, Rio Grande do Sul – Capital ou Rio Grande do Sul – Interior, o desconto médio caiu para 31%. Por outro lado, se a negociação tivesse ocorrido em outra região, subiu para 34%. No primeiro subconjunto ficaram 26% das observações, enquanto no segundo, 74%. Após, ambos os subconjuntos foram divididos pela velocidade (se maior que 1 metro por segundo<sup>32</sup>). A escolha por essa característica para determinar os pontos de corte nos dados é interessante, pois por mais que esteja se tratando de propostas negociadas em mercados diferentes, ela representa elevadores, em tese, semelhantes, evidenciando uma certa homogeneidade no produto demandado em mercados diferentes, com os clientes mais ou menos dispostos a pagar um prêmio na negociação a depender da região. Esse equipamento é definido pela companhia como uma obra do segmento de mercado classificado *low rise*<sup>33</sup>. Existem regiões em que ela sofre maior competição nesse mercado e, para vencer uma negociação, há maiores exigências financeiras, enquanto em outras localidades não há tanta pressão por preço. A partir das divisões de dados, esse mercado possui 75% das observações e, assim, pode-se inferir que tem uma relevância estratégica importante.

Os outros mercados além do *low rise* são o *mid rise* e o *high rise*. Identificar os dois últimos a partir das secções propostas no modelo não parece viável, pois precisar-se-ia de maiores especificidades quanto as características técnicas. No entanto, o que pareceu visível é que no mercado *low rise* parece se exigir menos desconto para vencer uma negociação, independente da região.

Com o modelo parametrizado para o cenário 1, o erro quadrático médio (RMSE)<sup>34</sup> foi de 8,4%, o erro médio absoluto (MAE)<sup>35</sup> de 5,9% e o erro médio relativo (MRE)<sup>36</sup> de

32 Aqui foi feita uma aproximação. O modelo considerou 75 metros por minuto como ponto de corte, no entanto, não existe equipamento com essa característica. As configurações válidas de velocidade nos produtos padrão são, em metros por minuto: 60, 90, 105, 120, 150, 180, 210 e 240. Podem ser observadas outras possibilidades em equipamentos especiais, como cargueiros ou montacargas. No entanto, são apenas 16 equipamentos dos 7.895 analisados.

33 Segmentação de características de equipamento a partir de velocidade e percurso. Ela foi implementada recentemente, não sendo possível classificar a base de equipamentos de maneira retroativa, pois é incerto se as premissas não teriam se alterado ao longo do tempo.

34 Medida estatística usada para avaliar a precisão de um modelo de previsão em relação aos dados reais. Determina a diferença entre as previsões feitas pelo modelo e os valores observados e, em seguida, calcula a média dos quadrados dessas diferenças – quanto menor o valor, mais preciso é o modelo. Essa métrica é especialmente utilizada para verificar o desempenho de modelos de previsão quando se trata de valores contínuos, como nesse caso, o desconto de negociações comerciais.

35 Medida que representa a média da diferença absoluta entre as previsões de um modelo e os valores reais. Na análise de regressão por *random forest*, é utilizada para avaliar a capacidade do modelo em prever a variável alvo com base nas variáveis selecionadas – quanto menor o valor, melhor é o modelo em suas previsões e mais próximas elas estão dos valores reais observados.

36 Métrica de precisão do modelo calculada a partir da média dos erros relativos (valores reais – valores previstos, multiplicados por 100%). Ao contrário do MAE, o MRE leva em consideração o tamanho dos valores reais, sendo útil para avaliar a precisão do modelo na predição de um conjunto de dados com valores extremos.

22,7%. Já para o segundo cenário, os mesmos indicadores foram, respectivamente, 9,0%, 6,5% e 25,0%. Já para o último, as métricas obtidas foram 2,9%, 2,1% e 6,7%, indicando uma performance superior aos que se utilizaram de outros filtros (tabela 1).

Tabela 1 - Métricas de análise dos cenários para o modelo de Árvore de Decisão

	<b>RMSE</b>	<b>MAE</b>	<b>MRE</b>
Cenário 1	8,4%	5,9%	22,7%
Cenário 2	8,9%	6,6%	21,7%
Cenário 3	2,9%	2,1%	6,5%

Fonte: o autor (2023).

De maneira geral, a performance dos dois primeiros cenários apresentou um erro quadrático médio bastante elevado quando testada a árvore de decisão para prever os valores de desconto obtidos na testagem. Em especial o cenário 2, que teve foi filtrado a partir da variável Modelo, apresentou um desempenho aquém do primeiro cenário, indicando que talvez tenha sido filtrado de maneira equivocada. Na análise por *random forest* poderá se obter resultados mais precisos, dada a natureza da técnica.

#### 4.2 *RANDOM FOREST*

Além da modelagem por árvore de decisão, também foi proposta a análise por *random forest*. Como visto, essa técnica consiste na execução de um amalgamado de árvores de decisão com um resultado médio entre as diferentes árvores como resultado. A implementação foi feita utilizando-se o pacote *randomforest*<sup>37</sup> no R.

Foram modelados os mesmos cenários que os estudados na árvore de decisão. O primeiro considerando apenas os filtros propostos no Capítulo 4, o segundo com exclusão dos modelos obsoletados, e o terceiro considerando propostas apenas a partir de 2016. Também

---

Uma baixa porcentagem indica que o modelo tem boa precisão em prever a variável alvo em função das variáveis independentes.

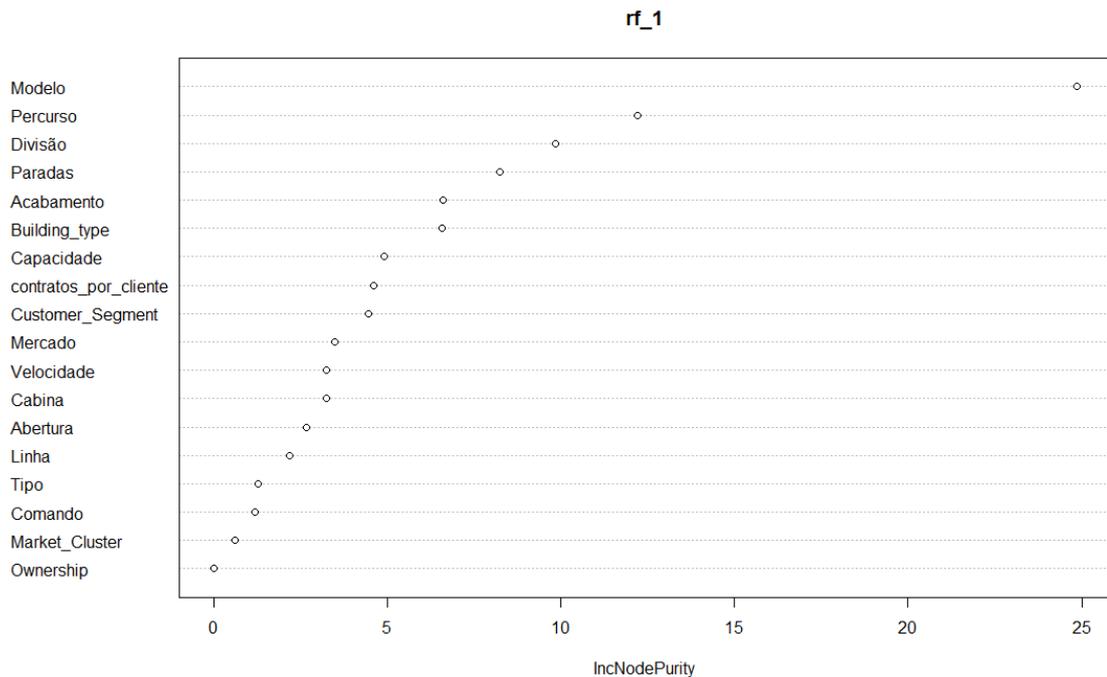
<sup>37</sup> O pacote "randomForest" é uma implementação em R do algoritmo de *random forest*. O algoritmo constrói  $n$  árvores de decisão durante o treinamento, usando amostras aleatórias dos dados de treinamento e características selecionadas aleatoriamente em cada divisão. A previsão é feita agregando as previsões de cada árvore. A função é altamente configurável e permite ajustar vários parâmetros, como número de árvores construídas e quantidade de nós em cada uma delas.

foram feitos testes de estresse para verificar a consistência do modelo, cuja metodologia e parametrização será exposta a seguir.

Os resultados de cada uma das análises foram similares aos obtidos na árvore de decisão única para cada cenário no que diz respeito à ordem de importância das variáveis determinantes do desconto (tabela). No *random forest* não há uma árvore de decisão única para analisar os nós, pois justamente o método consiste na agregação de uma multiplicidade de árvores para se obter o resultado, o que acaba limitando a visualização das tendências e secções dos conjuntos de dados como foi feito no subcapítulo 4.1.

Ainda assim, foi possível analisar as variáveis mais influentes na determinação do desconto para cada cenário. No primeiro, foram: Modelo (24,9), Percurso (12,2), Divisão (9,9), Paradas (8,2) e Acabamento (6,6).

Figura 4 - Ranking de importância das variáveis no cenário 1



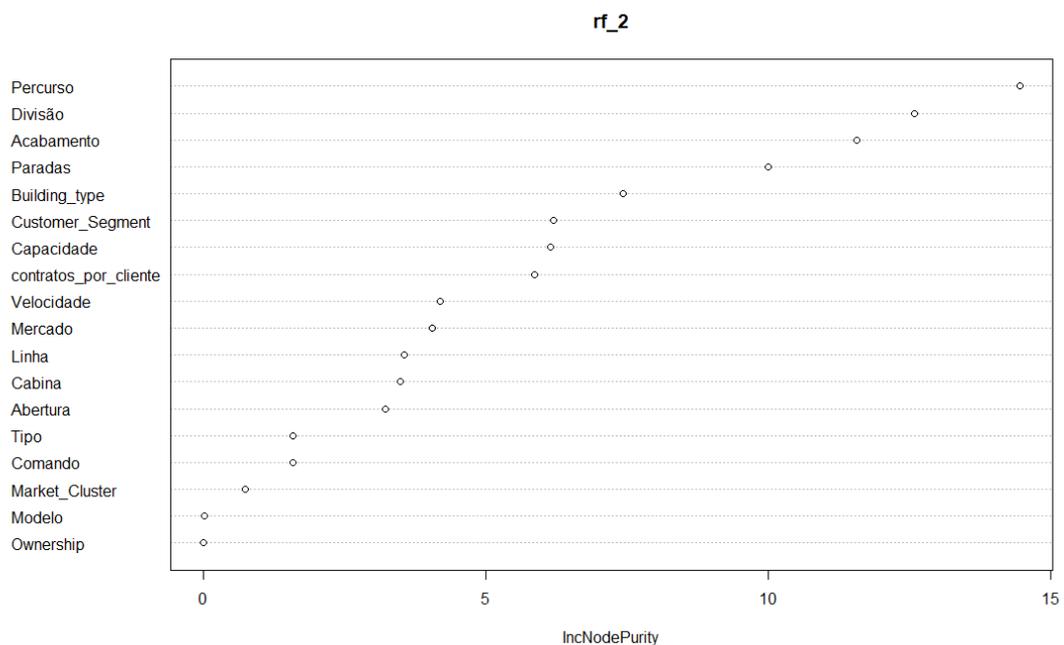
Fonte: o autor (2023).

Com o ajuste da variável Modelo em efeito apenas para os cenários 2 e 3, pode-se perceber uma similaridade com a priorização de importância observada na análise por árvore de decisão. Essa metodologia também deu ênfase a uma importante característica física de um

elevador na sua configuração: percurso. Outro fator que parece determinante, independentemente do cenário ou da técnica, é a abertura quanto às áreas administrativas geográficas em que a empresa atua, representada na variável Divisão.

A ordem de variáveis mais influentes para o modelo se alterou sensivelmente no cenário 2 quando comparado com o resultado obtido na avaliação por árvore de decisão: Percurso (14,5), Divisão (12,6), Acabamento (11,6), Paradas (10,0) e Building\_type (7,4). Aquele método havia trazido a preponderância dos atributos de Cabina, Modelo e Divisão. Em relação ao cenário 1, chama atenção a redução do peso da variável Cabina, enquanto Modelo, Abertura e Divisão permaneceram estáveis. Destaca-se aqui que, com os ajustes dos Modelos, a variável praticamente perde relevância na análise, pois a plataforma elétrica dos produtos da companhia não sofreu alguma alteração substancial, apenas uma cadeia sucessória de obsolescência à medida que um ou outro componente era homologado ou obsoleto. Após a normalização da variável, permaneceram apenas dois elementos, o Frequencedyne e o MHC2 – enquanto aquele é a plataforma básica para praticamente todos os equipamentos, o MHC2 tem aplicação mais específica, abrangendo equipamentos cuja especificação é bastante estrita em termos de normatização técnica e de mercado *mid rise*.

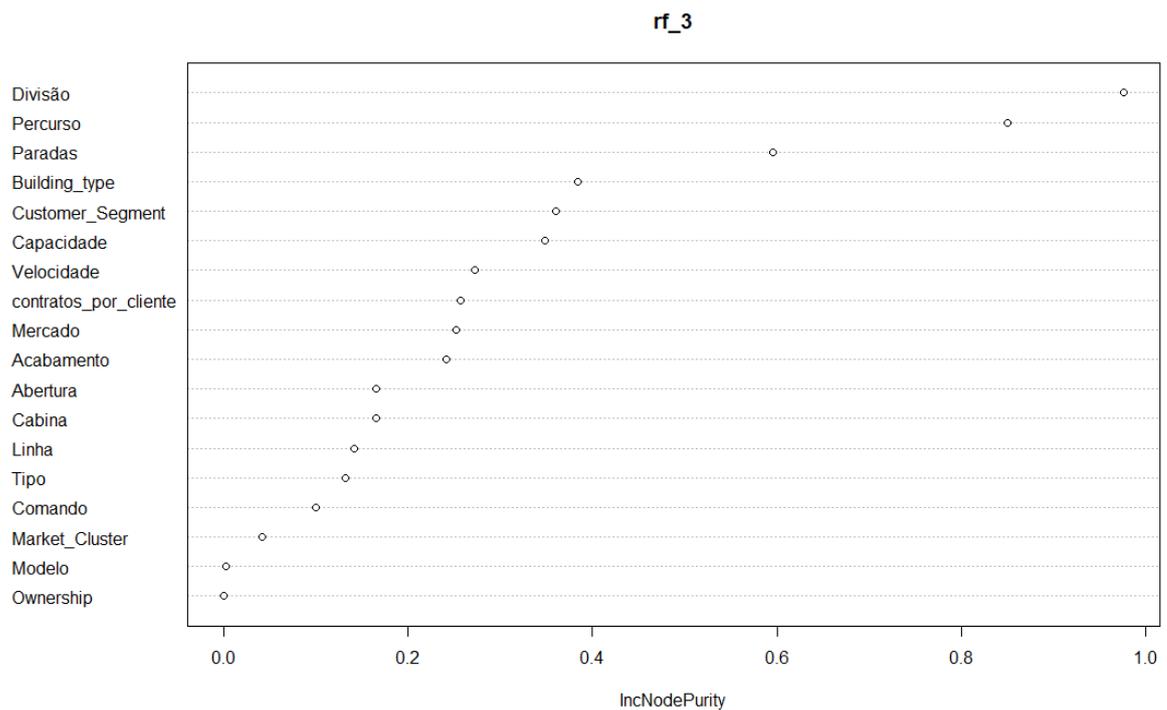
Figura 5 - Ranking de importância das variáveis no cenário 2



Fonte: o autor (2023).

O cenário 3 foi o que apresentou menor diferença entre o potencial de explicação de cada variável na classificação das variáveis mais influentes na predição do desconto. Merecem destaque Divisão (1,0), Percurso (0,9), Paradas (0,6), Building\_type (0,4), Customer\_segment (0,4) e Capacidade (0,3). Válido destacar que a variável Contratos\_por\_cliente em nenhum cenário apresentou-se como um fator de influência relevante na predição do desconto – essa variável foi criada para tentar avaliar a existência de figura de um cliente parceiro da empresa e foi definida simplesmente verificando quantos contratos cada cliente que contratou alguma proposta tinha na base, sem considerar se era o primeiro ou segundo contrato ou quantos contratos já se tinham negociado com aquele cliente no momento da negociação.

Figura 6 - Ranking de importância das variáveis no cenário 3



Fonte: o autor (2023).

Os três cenários modelados apresentaram resultados interessantes, coerentes com as variações entre as premissas estabelecidas, e, de maneira geral, como visto, as variáveis mais relevantes à predição da variável alvo foram similares ao apontado no estudo por árvore de decisão. O grupo de características mais definitivo na predição do desconto parece ser o que

agrupa atributos físicos do equipamento, como Cabina, Modelo, Percurso, Capacidade, Abertura ou Velocidade. Dados de mercado apareceram pontualmente e talvez em uma modelagem de cenários com intervalos menores possam aparecer outras tendências, posto que se esperaria que essas variáveis dependam mais de condições macroeconômicas de mercado do que propriamente de propriedades de um elevador contratado. Por exemplo, talvez a variável *Contratos\_por\_cliente* tenha ganhado relevância apenas no cenário que considera um menor intervalo temporal por uma eventual rotatividade da base de clientes ou mesmo dos participantes do mercado.

Afora os atributos físicos, o fator que pareceu preponderante foi a Divisão, independente do cenário ou do método. A árvore de decisão da imagem 3 também evidenciou que parece haver uma heterogeneidade maior entre a segmentação geográfica do mercado em que a companhia atua do que em relação às características dos equipamentos no que diz respeito à capacidade de explicação do desconto negociado.

Tabela 2 - Ranking de importância das variáveis

	<b>Cenário 1</b>	<b>Cenário 2</b>	<b>Cenário 3</b>
Abertura	2,7	3,2	0,2
Acabamento	6,6	11,6	0,2
Building_type	6,6	7,4	0,4
Cabina	3,2	3,5	0,2
Capacidade	4,9	6,1	0,3
Comando	1,2	1,6	0,1
contratos_por_cliente	4,6	5,9	0,3
Customer_Segment	4,5	6,2	0,4
Divisão	9,9	12,6	1,0
Linha	2,2	3,6	0,1
Market_Cluster	0,6	0,8	0,0
Mercado	3,5	4,1	0,3
Modelo	24,9	0,0	0,0

Ownership	0,0	0,0	0,0
Paradas	8,2	10,0	0,6
Percurso	12,2	14,5	0,9
Tipo	1,3	1,6	0,1
Velocidade	3,3	4,2	0,3

Fonte: o autor (2023).

O modelo em cada cenário foi testado com base em métricas de RMSE (erro quadrático médio), MAE (erro médio absoluto) e MRE (erro médio relativo). Após modelados os três cenários, foi realizado um teste de estresse no modelo para verificar a consistência das análises em cada um dos contextos. A metodologia utilizada foi idealizada utilizando-se um subconjunto dos dados para o treino do modelo e outro para o teste. Assim, utilizou-se as propostas pré-COVID-19 para tentar modelar o desconto das propostas do pós-pandemia, um período que além de marcante no cenário macroeconômico, coincide com o câmbio da política de desconto da companhia, o que deve impactar a capacidade de predição do modelo para cada cenário. Espera-se que no primeiro cenário, com um maior intervalo temporal considerado no conjunto de dados, o teste de estresse apresente uma capacidade de explicação inferior ao modelo original, com um erro quadrático médio maior, bem como MAE e MRE também maiores. Para o segundo cenário, dada a peculiaridade com que foi modelado, talvez se revele um erro maior percebido nas métricas de avaliação. Já para o terceiro quadro espera-se resultados bastante similares ao teste base.

Para o primeiro cenário testado, os resultados obtidos indicam que o modelo de teve um alto nível de precisão na previsão de medidas de desconto, com um RMSE de 5,8%, MAE de 3,4%, MRE de 14,2% e um EV de 66,4%. Comparando-os com os obtidos no teste de estresse (RMSE: 12,0%, MAE: 11,3%, MRE: 35,3%), percebe-se um aumento no erro dos resíduos obtidos ao se treinar o modelo com os dados do pré-pandemia e tentar predizer os valores de desconto do pós-pandemia a partir das configurações e caracterizações de equipamentos e propostas daquele período. A conclusão é coerente com a narrativa identificada acerca dos níveis de desconto, dado que o conjunto de dados analisados nesse cenário no que diz à variável alvo do modelo contempla os três períodos em que se variou a política de desconto (em que o desconto era conservador, o período em que era usado de maneira mais agressiva no mercado e quando volta a ser utilizado de maneira conservadora), já as métricas de avaliação entre os dois cenários testados aponta uma incidência muito maior

de erro nas predições. Isso se explica, além do argumento supracitado, por exemplo, pela diferença de características técnicas entre os equipamentos ao longo do tempo e pela evolução dos mercados de maneira heterogênea em diferentes regiões geográficas.

Já para o cenário 2, quando excluídos os modelos de equipamento que foram obsoletados e que representam um número significativo de observações entre os dois conjuntos de dados, percebe-se uma incidência maior de erros e uma queda no percentual de variação da variável alvo explicada pelo modelo: 62,5% em relação aos 66,4% anteriores, e o RMSE aumentou de 5,8% para 6,5%. Isso pode indicar que o filtro proposto no capítulo 4, que consistiu em simplesmente ajustar o conjunto de dados analisados para as plataformas elétricas atualmente produzidas pela manufatura da empresa, possa não ter sido o mais adequado no tratamento dos dados. O MAE e o MRE variaram, respectivamente, de 3,4% para 4,3% e de 14,2% para 14,1%. No entanto, quando se compara com o teste de estresse, o RMSE foi de 4,4%, o MAE de 3,4%, o MRE foi 10,7% e a variação explicada pelo modelo foi de 66,2%. Não apenas o erro na predição caiu, como a variação explicada pelo modelo aumentou.

Testando o modelo para o cenário 3, percebe-se maior precisão em relação aos outros resultados apresentados, com um RMSE de 2,2%. Com esses parâmetros, a proporção da variação explicada foi apenas de 56,2%, frente aos 66,4% e aos 62,5% dos cenários 1 e 2, respectivamente. O MAE e o MRE apurados foram de 1,4% e 4,5%, indicando que houve uma variância menor em relação aos outros contextos testados. Esse resultado é esperado, dada a redução do ruído ocasionada por uma amostragem mais enxuta, considerando apenas propostas fechadas a partir de 2016, e mais em linha com as políticas de desconto atuais. O teste de estresse revelou resultados em linha com os obtidos na avaliação base, com indicadores de erro levemente acima (RMSE: 3,0%, MAE: 2,3% e MRE: 7,3%). De todos os cenários avaliados, esse foi o que o teste de estresse representou um menor descolamento em relação às premissas de negócio provadas, mas menor capacidade de predição da variável alvo.

Tabela 3 - Métricas de análise dos cenários para o modelo de *Random forest*

	<b>RMSE</b>	<b>MAE</b>	<b>MRE</b>
Cenário 1	5,8%	3,4%	14,2%
Cenário 1 - Teste de estresse	12,0%	11,3%	35,3%

Cenário 2	6,5%	4,3%	14,1%
Cenário 2 - Teste de estresse	4,4%	3,4%	10,7%
Cenário 3	2,2%	1,4%	4,5%
Cenário 3 - Teste de estresse	3,0%	2,3%	7,3%

Fonte: o autor (2023).

De maneira geral, dentre os três cenários modelados, o que teve melhor desempenho foi o terceiro. O segundo, como observado, pôde ter sido mal avaliado com o filtro por modelo, e isso refletiu num erro maior do modelo e uma capacidade de predição inferior aos outros dois. Já o primeiro, foi consistente, mas aquém dos resultados obtidos no último contexto, dado que foi filtrado de maneira a excluir uma parcela do conjunto de dados que não reflete bem a atual estratégia de precificação de contratos da empresa.

As métricas reportadas no teste de estresse do cenário 1 surpreenderam, com uma capacidade de predição da variável alvo superior ao da avaliação original com a mesma parametrização de filtros e seleções. Nesse contexto, a proporção de variação da variável alvo explicada pelo modelo foi de 66,4%, enquanto no teste de estresse foi de 70,9%. Um dos motivos que pode apresentar uma explicação para esse panorama é um *overfitting* da amostra de teste – fenômeno usual quando se modela uma base de dados com muitas dimensões, o que resulta num modelo demasiado complexo. As métricas de erro, entretanto, apontaram a um desempenho inferior. Comparando-o ao cenário 2, teve um desempenho bem melhor, tanto em função da capacidade de predição como do menor erro. Este contexto, ainda quando avaliado no teste de estresse, apresentou um resultado ainda mais incondizente com os resultados esperados, com métricas de erros menores e maior capacidade preditiva no modelo de estresse.

Já o cenário 3, apresentou os melhores indicadores de erro, mas menor capacidade preditiva do modelo. O teste de estresse apresentou melhor desempenho que o de outros cenários e ficou em linha com os resultados obtidos na avaliação base.

#### 4.3 *EXTREME GRADIENT BOOSTING*

A partir das várias análises realizadas até agora, o cenário 3 parece o mais consistente em termos de capacidade de predição e com resultados mais adequados. A fim de aprofundar o seu estudo e apresentar uma tentativa de solução para o problema de *overfitting*

no modelo, foi trabalhado um modelo de *extreme gradient boosting* a partir da implementação proposta pelo pacote *xgBoost* no R.

O *XGBoost* é uma técnica baseada em árvores de decisão amplamente utilizada em problemas de classificação e regressão. Baseia-se em uma construção sequencial de árvores de decisão e utiliza regularização de coluna<sup>38</sup> para prevenir *overfitting* e melhorar a capacidade de generalização do modelo.

A construção sequencial de árvores de decisão no *XGBoost* começa com a criação de uma árvore simples e, em seguida, adiciona novas árvores que buscam corrigir os erros da árvore anterior. Essa técnica é conhecida como *boosting*<sup>39</sup> e é uma das principais inovações do *XGBoost* em relação ao *random forest*, que constrói um conjunto de árvores independentes simultaneamente.

Em que pese as vantagens em relação aos outros modelos, a preparação do conjunto de dados é mais trabalhosa, necessitando-se *dumminizar* as variáveis categóricas em numéricas. Assim, cada coluna da matriz foi aberta em tantas colunas quantas forem necessárias para descrever a totalidade de variáveis diferentes (por exemplo, a coluna Linha foi aberta em Linha – FDN, Linha – FDG, linha Synergy 100...), e a proposta seria então identificada por um fator binário para cada propriedade. Assim, as dimensões do conjunto de dados aumentaram de 20 para 101.

Essa etapa de tratamento dos dados, além de ser passo exigido pelo método, também permite a abertura das variáveis para terminar o quê, dentre elas, é de fato mais relevante. O resultado permite identificar, por exemplo se a Divisão for um fator relevante na predição do desconto, quais as divisões que apresentam maior sensibilidade à variável alvo.

A primeira abertura dos dados categóricos que pode ser observada foi em relação à divisão, em que a Divisão – Rio Grande do Sul Interior foi determinada como a mais sensível ao modelo, em seguida Divisão – São Paulo Capital. Essa análise permite concluir que de fato a empresa atua num mercado heterogêneo, em que cada região tem suas características e fatores determinantes para um maior ou menor desconto no fechamento de negociações. A região interior do Rio Grande do Sul é tida como uma espécie de mercado cativo, tanto pela questão de que a empresa analisada tem suas origens no estado gaúcho e há um apelo

---

38 É uma técnica voltada para melhorar a capacidade de generalização do modelo prevenindo o *overfitting*. Com a sua implementação, o modelo limita o número de propriedades testadas em cada árvore, evitando a criação de árvores complexas e que se ajustem demasiadamente aos dados de treinamento, apresentando, assim, baixa capacidade de generalização.

39 Esse método é capaz de melhorar a precisão das previsões em modelos complexos e com dados de alta dimensionalidade através de um processo sequencial que consiste em treinar um modelo fraco em todo o conjunto de dados e, em seguida, novos modelos ajustados para as previsões incorretas em cada iteração.

identitário para o cliente, como a proximidade da planta fabril em relação às principais concorrentes confere uma vantagem de tempo e valor entrega. Já a região de São Paulo – Capital tem algumas características que a destacam de outras, como a altura das edificações construídas costumam ser mais alta que em outras regiões, o acabamento de obras usualmente exigir um produto mais customizado e existe maior concorrência de outras empresas.

A despeito de a Divisão ser um importante fator de análise no desconto de uma proposta, como conferido pelos modelos, o fator mais influente foi o Percurso. Após, Divisão – Rio Grande do Sul Interior, Paradas, Velocidade e Divisão – São Paulo Capital e Capacidade. Após, o `Contratos_por_cliente` foi a variável mais influente no desconto.

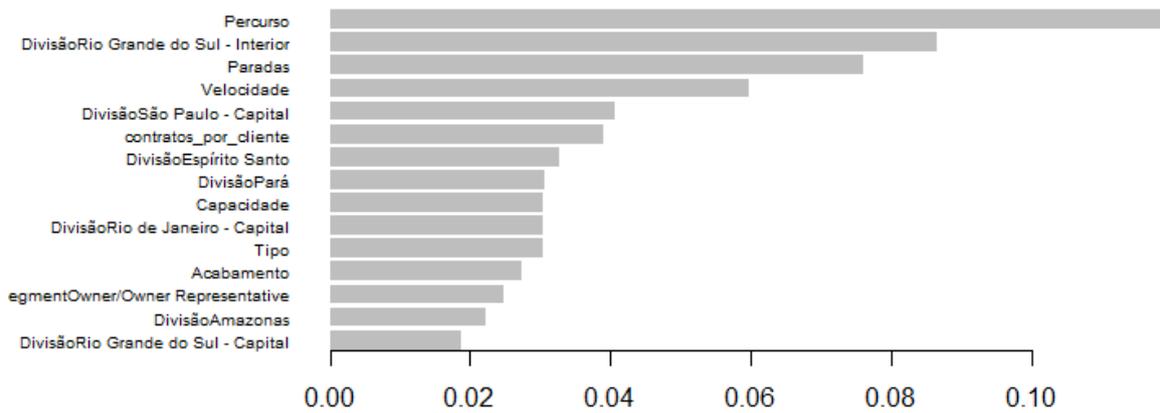
As propriedades determinantes para a variável alvo no teste de árvore de decisão para o cenário 3 foram preponderantemente respectivas às propriedades físicas do elevador. Quando modelado pelo *random forest*, percebeu-se uma influência também do contexto de mercado como fatores decisivos para a predição do desconto de uma negociação. Já quando o *gradient boosting* foi testado nesse contexto, pôde-se apurar que não só as características físicas do equipamento são fundamentais, mas também o contexto de mercado em que a negociação está sendo feito.

Interessante notar que os fatores físicos mais influentes (Percurso e Velocidade) são os que determinam a classificação *low rise*, *mid rise* e *high rise*, visto no capítulo 4.1. Já a Divisão é uma dimensão bastante heterogênea na capacidade de predição do desconto: se Rio Grande do Sul Interior e São Paulo Capital são Divisões particulares o suficiente para terem uma consolidação de mercado por si, Divisões como Goiás e Paraná Interior são praticamente irrelevantes na determinação da variável alvo. Outro bom exemplo desse fenômeno ocorreu com a variável Linha: Linhas que determinam um intervalo mais estrito de configuração de equipamento apresentaram maior capacidade preditiva, enquanto as que possibilitam maior customização mostraram menor relevância (respectivamente, pode-se identificar o Synergy 100 S e o FDN).

Também algumas combinações de variáveis que são vinculadas na base de dados apresentaram similar poder de explicação, como poderia se esperar (a Linha Evolution ficou muito próximo à Cabina P4000). Entretanto, um exemplo interessante de se analisar é referente as Linhas Synergy 100 e Synergy 100S, que têm como única Cabina possível a KM18, mas enquanto essa apresentou um poder de predição muito próximo à Cabina, aquela Linha ficou bastante distante. A diferença entre esses produtos é que, como já visto, o último é apresentado ao cliente como uma versão mais restrita do primeiro, em uma tentativa de padronizar as linhas de produto da empresa, e, para tanto, é oferecido um desconto de 7%

implícito no preço da proposta. Além da abertura por Divisão, a visualização de como cada uma das Linhas pode prever melhor ou pior o desconto é um descobrimento muito importante, dado que é por como o portfólio de produtos hoje é dividido.

Figura 7 - Ranking de importância das variáveis no cenário 3



Fonte: o autor (2023).

O *ranking* de variáveis apresentado parece coerente com as conclusões extraídas no capítulo 4.1, que apontaram que elevadores que se inseririam na classificação *low rise* exigiriam menor desconto para concluir uma negociação. Os fatores que foram mais determinantes para o desconto nessa análise foram justamente características trancadas na parametrização de um Synergy 100 S, equipamento voltado para a aplicação *low rise* e que foi a mais influente no modelo. Entretanto, isso pode significar que há um ruído na base de dados em virtude do desconto implícito no custo dado a essa linha, de maneira a enviesar as considerações de todas as outras variáveis.

Quanto às métricas de desempenho, o modelo foi testado num método chamado *cross-validation*<sup>40</sup> para avaliar questões como erro e *overfitting* dos dados. Nessa análise, o

<sup>40</sup> O objetivo do teste é avaliar a habilidade do modelo de fazer previsões precisas em dados não vistos, que não foram usados para treinar o modelo. Na aplicação dessa técnica, o conjunto de dados disponível é dividido em dois ou mais subconjuntos, e um subconjunto é usado para treinar o modelo e o outro é usado para testar o modelo. O processo é repetido  $n$  vezes, com diferentes subconjuntos de treinamento e teste, para que o modelo seja testado em diferentes conjuntos de dados. A forma de avaliação utilizada foi a *k-fold validation*, que consiste em dividir a amostra em  $k$  subconjuntos iguais e testar cada vez uma parte diferente como conjunto de validação

modelo apresentou um RMSE de 2,0% para a amostragem de treino e 2,5% para a de teste, indicando que realmente existe *overfitting* dos dados. O MAE percebido foi de 1,5% na amostragem de treino e 1,8% na de teste e o MRE, respectivamente, 4,6% e 5,5%.

De maneira geral, as métricas de erro apontam para um desempenho do modelo similar ao obtido no *random forest*. Da mesma forma, o *ranking* de importância das variáveis obtidas na análise de *gradient boosting* também possibilita inferir um resultado similar àquele modelo, com ênfase na relevância de características físicas e de geografia de mercado. A grande contribuição que essa análise conferiu às conclusões do estudo foi a possibilidade de abertura das variáveis e permitir ver certa heterogeneidade na capacidade de previsão de diferentes propriedades de uma mesma variável (Divisão e Linha foram bons exemplos).

Tabela 4 - Métricas de análise do cenário 3 para cada modelo testado

	<b>RMSE</b>	<b>MAE</b>	<b>MRE</b>
Árvore de decisão	2,9%	2,1%	6,5%
<i>Random forest</i>	2,2%	1,4%	4,5%
<i>Gradient boosting</i>	2,4%	1,7%	5,3%

Fonte: o autor (2023).

## 5 RESULTADOS

Nesse estudo visou-se identificar que fatores podem influenciar na concessão de desconto e precificação de novos contratos de compra e venda de elevadores. A base analisada contemplou todas as vendas efetivadas de uma indústria manufatureira do ramo entre os anos de 2012 e 2022. O conjunto de dados disponibilizado contém, dentre características físicas, segmentações de mercado e informações financeiras, 195 variáveis e 15.279 observações.

Para identificar quais fatores são influentes na variável alvo, conforme a melhor literatura, exposta no capítulo 2, foi realizada uma análise por árvore de decisão, em que foram testados 3 cenários distintos. O primeiro contemplou a totalidade das observações, mas o modelo apresentou métricas de desempenho insatisfatórias e baixa capacidade de predição, e, assim, optou-se por realizar uma segmentação no conjunto de dados e testar um novo cenário. A partir dessa nova análise, pôde-se verificar um modelo com piores capacidades de explicação e métricas de performance – novamente optou-se por uma segmentação de dados, em que se excluiu todas as negociações anteriores a 2015 do conjunto. O terceiro cenário apresentou boas métricas de desempenho e capacidade de determinação da variável alvo.

A seguir, os cenários foram testados em um modelo de *random forest*, com parametrização idêntica para cada um, apenas cambiando-se o *input* de dados. Os resultados foram semelhantes, nos três contextos, aos obtidos na análise por árvore de decisão, bem como as métricas de desempenho do modelo em cada contexto foram similares às obtidas naquele modelo.

Após, foi realizado um teste de estresse no modelo para cada cenário, em que os dados foram segmentados a partir do marco da pandemia da COVID-19, e propostas anteriores foram agrupadas para o treino do modelo e tentar prever o desconto das negociações do pós-pandemia. Essa referência foi escolhida em virtude tanto das alterações no cenário macroeconômico incorridos em virtude das medidas preventivas à propagação do vírus, bem como pela mudança na orientação do desconto nas vendas da empresa estudada. Os resultados obtidos foram coerentes com os do modelo em cada cenário.

Por fim, identificando o cenário 3 como o mais adequado nas análises, optou-se por aprofundar o estudo com uma modelagem de *gradient boosting*. Essa técnica permite visualizar a influência de quais propriedades dentre as variáveis categóricas tem capacidade de predição da variável alvo. Também foi realizado um teste de estresse por *k-fold cross-validation*, em que pôde ser verificado que o conjunto de dados apresentou *overfitting* no

testagem do modelo – entretanto, o grau de *overfitting* foi moderado, de maneira a não prejudicar as conclusões extraídas.

Das análises realizadas, foi possível inferir que tanto características físicas quanto de mercado são importantes na avaliação do desconto de uma proposta comercial. A partir da construção da árvore de decisão, parece possível concluir que as propriedades que definem um equipamento como um *low rise*, *mid rise* ou *high rise* são as mais críticas dentre as que tangem o equipamento em si – especificamente, a partir da árvore de decisão construída no capítulo 4.1, é possível identificar que, de maneira geral, os elevadores *low rise* demandam menos desconto para serem negociados do que seus pares *mid rise* e *high rise*.

Isso também tem impacto na heterogeneidade de algumas variáveis, como Divisão, dado que esse grupo representa regiões geográficas distintas e mercados específicos, o que foi possível identificar na análise de *gradient boosting*. Mercados onde a empresa enxerga menor competição também têm características de prédios de menor Percurso – então é natural de se esperar que as negociações irão naturalmente demandar menos desconto e, em sua maioria, serão demandados equipamentos *low rise*. Por outro lado, uma Divisão como São Paulo Capital, naturalmente demanda da companhia elevadores mais customizados, com maiores Percurso, Velocidade e Capacidade, ao passo que a região encara uma concorrência maior de competidores.

Outra questão relevante percebida foi, com a implementação de Linhas com desconto implícito para produtos padronizados, em virtude de iniciativas da empresa de simplificar o portfólio de produtos e obter ganhos de escala com um grande volume de fabricação, a possibilidade de se enxergar alguns vieses na base de dados. Isso pôde ser notado no cenário 1 em que a árvore de decisão construída, em seu primeiro nó, dividiu a amostra a partir exatamente da Cabina One, a única possível para o produto comercial Synergy One, que demandava em média 19% de desconto, enquanto o resto dos elevadores demandaria 34%. O mesmo efeito pôde ser notado no cenário 3, em que a análise de *gradient boosting* evidenciou que, dentre a variável Linha, a propriedade mais importante era aquela relativa ao Synergy 100 S, outro expoente dessa concepção de padronização.

Uma tendência percebida foi como a alteração da política de desconto da companhia ao longo do tempo afetou a variável alvo. No cenário 1, a média do desconto foi de 37% e a correlação com a variável Ano foi -0,46, já no cenário 2, as medições foram, respectivamente, 35% e -0,29, e, no cenário 3, 33% e 0,16. Outro ponto importante para a concessão de descontos, é relativo às Divisões (utilizando-se apenas o cenário 3 como referência, a diferença de desconto entre São Paulo – Capital e Rio Grande do Sul – Interior já é visível:

respectivamente, 39% e 33%), pois na estrutura organizacional da companhia, cada divisão funciona praticamente como uma empresa *standalone*, com o gerente tendo grande autonomia nas decisões de negócio – ainda estão sujeitas às políticas corporativas, mas diferentes metas de crescimento podem impulsionar a gestão para caminhos diversos.

A análise da variável *Contratos\_por\_cliente* efetivamente não possibilitou confirmar a suspeita de a relação da empresa ser diferente quando se trata de um cliente com mais contratos já negociados do que com clientes que fecharam apenas uma proposta. A variável definida foi uma síntese bastante simplificada do argumento, sendo, talvez, insuficiente para a análise. Parece, assim, claro que também dever-se-ia investigar outras relações, como se o desconto tende a reduzir à medida que novos acordos são negociados.

Em conclusão, pôde-se identificar que o Desconto não é homogêneo ao longo do tempo, na cobertura geográfica do mercado brasileiro, por produto ofertado ou por segmento e tipo de cliente. Mesmo com as ressalvas feitas em relação aos vieses percebidos, a companhia parece ter um mercado consolidado de elevadores com características de baixa Capacidade, Velocidade e Percurso, mas faltar com um produto competitivo para segmentos que exigem um equipamento mais customizado. Também ficou claro que regiões diferentes podem demandar estratégias e produtos díspares, bem como apresentar uma posição já consolidada ou em disputa com a concorrência.

Para ilustrar as conclusões extraídas, exemplificamos uma amostragem de alguns equipamentos, com suas características e o desconto real comparado com a predição. A partir da sua análise, pode-se verificar alguns dos argumentos estudados.

Tabela 5 – Amostragens da predição de desconto no cenário 3

<b>Equip.</b>	<b>Divisão</b>	<b>Linha</b>	<b>Cabina<sup>41</sup></b>	<b>Vel.</b>	<b>Cap.</b>	<b>Par.</b>	<b>Per.</b>	<b>Desc.</b>	<b>Pred.</b>
1	CE	FDN	NAM	60	600	4	13	32,2%	31,8%
2	RS - Interior	Synergy 100	KM18	60	600	7	21	32,5%	31,2%
3	CE	FDN	NAM	105	750	23	64	39,8%	36,3%
4	SP - Capital	Synergy 100	KM18	60	600	8	21	33,1%	33,2%
5	PR - Capital	FDN	NAC	60	600	12	39	34,8%	34,2%
6	RS - Interior	Synergy 100	KM18	60	450	2	8	30,0%	31,6%
7	RS - Interior	Synergy 100	KM18	60	450	5	12	30,6%	30,5%
8	SP - Capital	FDN	NAM	90	600	11	33	35,8%	33,9%
9	PE	FDN	NAM	60	1200	4	12	38,6%	37,4%
10	SC	Synergy 100	KM18	60	600	7	19	32,5%	33,5%
11	MG - Capital	Synergy 100S	KM18	60	600	10	29	30,2%	31,9%
12	ES	FDN	NAM	105	675	12	32	33,6%	33,8%
13	SC	FDN	NXP	105	750	16	52	36,5%	35,3%
14	SP - Capital	FDN	NAM	105	900	25	84	37,5%	36,9%
15	RS - Interior	FDN	NAM	60	1200	2	4	34,5%	33,1%

Fonte: o autor (2023).

---

41 As cabinas New Amazonm, New Art Collection e New Export foram, respectivamente, abreviadas como NAM, NAC e NXP.

## 6 CONSIDERAÇÕES FINAIS

A partir das análises realizadas, em relação aos objetivos específicos, atingiu-se de maneira satisfatória o que tange a investigar os vetores do desconto nas propostas comerciais e as tendências de comportamento das variáveis que o afetam. Já no que toca à verificação da figura do cliente parceiro, a pesquisa não logrou êxito em identificar, a partir da base de dados, se ela existe e suas possíveis características.

O desenvolvimento dessa pesquisa, estruturada de acordo com melhores práticas da indústria, esbarrou em alguns ajustes que foram necessários ao longo da pesquisa, mas que servem de lição para o tratamento gerencial de alguns fatores na companhia (por exemplo, como considerar a sucessão de um produto por outro quando se considera questões comerciais). A estruturação de dois modelos de análise foi satisfatória e ficou evidente as diferenças conceituais entre eles. Por fim, pôde-se verificar, por uma análise de *cross-validation* que os modelos apresentaram *overfitting*, o que pode ser corrigido pensando-se em simplificações nas variáveis consideradas, reduzindo-se o nível de granularidade ou propondo-se ajustes.

As conclusões obtidas nessa pesquisa ajudarão a validar metodologias quantitativas pelas quais a empresa poderá implementar ferramentas de formulação de desconto de referência em propostas comerciais de obras novas. As áreas corporativas, em tarefas como o *budgeting*, poderão ter visibilidade em determinar, para cada Divisão, quais os *drivers* de crescimento desejados de acordo com a estratégia do corpo diretivo da empresa.

Para aprofundar as conclusões, sugere-se aglutinar algumas características técnicas em *proxies*, como por exemplo a nomenclatura *low/mid/high rise*. Também, seria frutífero investigar como se comportariam as análises com alguns elementos congelados: como se comportariam os resultados considerando-se apenas a Divisão Rio Grande do Sul – Interior, ou São Paulo – Capital. Especialmente esse ponto pode ajudar a desenvolver soluções para os *gaps* de mercado que a empresa encontra na relação com sua base de clientes. Por fim, em relação ao cliente parceiro, deveria ser proposta outra forma de análise da variável, a partir de premissas mais bem definidas.

## REFERÊNCIAS

ABDELNOUR, A.; BABBITZ, T.; MOSS, S. **Pricing through the pandemic**: Getting ready for recovery. Building new pricing discipline, flexibility, and capabilities now can create long-term competitive advantages. [s.l.]: McKinsey & Company, 2020.

ADENTUJI, Abigail Bola *et alia*. House Price Prediction using Random Forest Machine Learning Technique. **Procedia Computer Science**, vol. 199, p. 806-813, 2022.

BONACCHI, Massimiliano; PEREGO, Paolo. Customer Analytics: Definitions, Measurement and Models. **SpringerBriefs in Accounting**, [s. l.], nov. 2018. Disponível em: [https://doi.org/10.1007/978-3-030-01971-6\\_2](https://doi.org/10.1007/978-3-030-01971-6_2). Acesso em: 16 mar. 2023.

BUDAK, Aysenur; SARVARI, Peiman. Profit margin prediction in sustainable road freight transportation using machine learning. **Journal of Cleaner Production**, [s. l.], v. 1, ed. 314, p. 1-19, jun. 2021. Disponível em: <https://doi.org/10.1016/j.jclepro.2021.127990>. Acesso em: 16 mar. 2023.

GERMANN, Frank; LILIEN, Gary; FIEDLER, Lars; KRAUS, Matthias. Do Retailers Benefit from Deploying Customer Analytics?. **Journal of Retailing**, [s. l.], v. 90, n. 4, p. 587-593, dez. 2014. Disponível em: <https://doi.org/10.1016/j.jretai.2014.08.002>. Acesso em: 16 mar. 2023.

INSTITUTE OF MANAGEMENT ACCOUNTANTS. **Developing an Effective Managerial Costing Model**: Statement on Management Accounting. Montvale: IMA, 2019.

LARIVIÈRE, Bart; VAN DEN POEL, Dirk. Predicting customer retention and profitability by using random forests and regression forests techniques. **Expert Systems with Applications**, [s. l.], v. 29, n. 2, p. 472-484, ago. 2005. Disponível em: <https://doi.org/10.1016/j.eswa.2005.04.043>. Acesso em: 16 mar. 2023.

LEE, Chee Sun; CHEANG, Peck Yeng Sharon; MOSLEHPOUR, Massoud. Predictive Analytics in Business Analytics: Decision Tree. **Advances in Decision Sciences**, v. 26, p. 1-30, mar. 2022.

LIU, Chang *et alia*. Forecasting copper prices by decision tree learning. **Resources Policy**, v. 52, p. 427-434, 2017.

LUCE, Fernando Bins; SILVEIRA, Cleo Schmitt; HELDT, Rodrigo. Predicting customer value per product: From RFM to RFM/P. **Journal of Business Research**, [s. l.], v. 127, p. 444-453, abr. 2021. Disponível em: <https://doi.org/10.1016/j.jbusres.2019.05.001>. Acesso em: 21 set. 2022.

MCKINSEY DIGITAL. **STRATEGY & CORPORATE FINANCE PRACTICES**: How COVID-19 has pushed companies over the technology tipping point—and transformed business forever. A new survey finds that responses to COVID-19 have speeded the adoption of digital technologies by several years—and that many of these changes could be here for the long haul. [s.l.]: McKinsey, 2020. Disponível em: <https://www.mckinsey.com/business->

functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever. Acesso em: 28 fev 2023.

PRAKASH, David; SPANN, Martin. Dynamic pricing and reference price effects. **Journal of Business Research**, vol. 152, p. 300-314, 2022.

PRICING Strategies. **AccountingTools**, 2022. Disponível em: <<https://www.accountingtools.com/articles/pricing-strategies>>. Acesso em: 16 mar. 2023.

RATHAN, Karunya; SAI, Somarouthu Venkat; MANIKANTA, Tubati Sai. Crypto-Currency price prediction using Decision Tree and Regression techniques. **Proceedings of the Third International Conference on Trends in Electronics and Informatics**. [S.l.]: IEEE, 2019.

ROSS, Stephen A.; WESTERFIELD, Randolph W.; JAFFE, Jeffrey. Administração financeira. 10. ed. São Paulo: AMGH, 2015.

SILVEIRA, Cleo Schmitt; OLIVEIRA, Marthaa Olivia Rovedder de; LUCE, Fernando Bins. Customer equity and market value: Two methods, same results?. **Journal of Business Research**, [s. l.], v. 65, n. 12, p. 1752-1758, dez. 2012. Disponível em: <https://doi.org/10.1016/j.jbusres.2011.10.034>. Acesso em: 16 mar. 2023.

WU, Meng; RAN, Yun; ZHU, Stuart X. Optimal pricing strategy: How to sell to strategic consumers? **International Journal of Production Economics**, v. 244, p. 1-16, 2022.

WIRTH, Rudiger; HIPPE, Jochen. CRISP-DM: Towards a Standard Process Model for Data Mining. **Practical application of knowledge discovery and data mining**, p. 29-40, 2000.

YE, Taogeng; YANG, Huiqiang, Price and Quality Management with Strategic Consumers: Whether to Introduce a High or Low Product Variant. **Applied Mathematics and Computation**, v. 386, p. 1-16, 2020.

YIU, Tony. Understanding Random Forest. **Towards Data Science**, 2019. Disponível em: <<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>>. 16 mar. 2023.

ZHA, Yong *et alia*. Credit offering strategy and dynamic pricing in the presence of consumer strategic behavior. **European Journal of Operational Research**, v. 303, p. 753-766, 2022.

INCLUIR UM PARAGRAFO NAS CONCLUSÕES EXPLICITANDO A  
APLICABILIDADE DO TRABALHO